Report on an Experimental Use of the DELPHI Technique

Office of Planning, Programming, & Budgeting

February 1972

Approved For Release 2005/04/18 : CIA-RDP82M00531R000400200002-1

Report on an Experimental Use of the DELPHI Technique

[redacted]

Planning & Research Staff

Office of Planning, Programming, & Budgeting

February 1972

25X1

PREFACE

This report describes the initial study on the use of the so-called Delphi technique by the Office of Planning, Programming and Budgeting. As such, it is primarily an illustration of method and does not, under any circumstances, represent an applied exercise with substantive results suitable for policy consideration by decision-makers.

No claims are made, or can be made, for the reliability of the predictions or evaluations because of the experimental nature of the exercise, the arbitrary method of selecting respondents and the subject matter.

[redacted]
Planning & Research Staff
O/PPB

25X1

Taking a look at the kinds of information that can play a role in decision-making, there are roughly three types. On the one hand, there are assertions that are highly confirmed--assertions for which there is a great deal of evidence backing them up. This kind of information can be called knowledge. At the other end of the scale is material that has little or no evidential backing. Such material is usually called speculation. In between is a broad area of material for which there is some basis for belief but that is not sufficiently confirmed to warrant being called knowledge. There is no good name for this middling area. I call it opinion. The dividing lines between these three are very fuzzy, and the gross trichotomy smears over the large differences that exist within types. However, the three-way split has many advantages over the more common tendency to dismiss whatever is not knowledge as mere speculation.

Where in this scale do the products of judgment, wisdom, insight, and similar intellectual processes lie? Not in speculation, we hope. And, almost by definition, not in knowledge. The most reasonable interpretation would be that these are flattering names for kinds of opinion. One might say, "Wisdom is opinion with charisma."

25X1

SUMMARY

The Delphi technique is a method of eliciting and refining group judgments, or opinions, where exact knowledge is not available. The procedures have three features: (1) Anonymous response and debate--opinions of respondents are obtained by formal questionnaires. (2) Iteration and controlled feedback--interaction is effected by a systematic exercise conducted in several iterations (rounds), with information feedback from round to round. (3) Statistical group response--the group opinion is defined as the aggregate of individual opinions on the final round and expressed in terms of two statistical indices. These features are designed to minimize the biasing effects of dominant individuals, or irrelevant communications and group pressure toward conformity [4]. Instead of using the traditional approach toward achieving consensus through open discussion, e.g., committees and conferences, Delphi eliminates committee activity altogether; thus reducing the influence of the foregoing psychological factors.

This report describes the initial study on the use of the Delphi technique by the Office of Planning, Programming and Budgeting. In the summer of 1970, fifteen respondents (volunteers) from several directorates and career services participated in the three-round exercise which examined the subject of Career Management.

Results of the exercise illuminate a number of points: the contents of the answers themselves, the basis on which respondents claimed their answers were made, the spread of "expert" views, the convergence of views following information and data feedback, the "experts" critique of each other's views, and not least of all, the shortcomings of the initial design and the suggested means for improving it. It does appear that some of the observed or suspected defects in the design, particularly the measurement of narrative-evaluative questions, can be eliminated on the basis of what has been learned from this experiment.

Four key and encouraging factors emerged from this exercise, viz., (a) the scope of Delphi is much broader than previously thought, therefore; (b) applications are not limited to technological probes of the future--in fact--Delphi can be used in any context where it is appropriate to seek a consensus among experts on a

particular issue; (c) the use of anonymous questionnaires generated a candor of responses that exceeded all expectations and; (d) care must be exercised on several technical factors to insure a respectable design and product.

ACKNOWLEDGMENTS

We wish to express our appreciation to Messrs. [redacted] Charles Briggs and [redacted] for their willingness to investigate Delphi beyond the discussion stages and sanction this pilot study. It is, also, a pleasure to thank [redacted] for his many helpful suggestions and improvements in this experimental exercise. We are indebted to Mr. Briggs for the additional time to write this report--now a year and half after the exercise--rather than let the findings mellow in some file and eventually make their way to the burn bag.

Some explanation concerning the delay of this report is in order. At the conclusion of the exercise the requests for briefings were at such a level that little time was available to document this exercise. We don't know if these requests reflected a keen interest or just a curiosity in Delphi but, in either case, time to document was not in the schedule. Then too, we had several takers for new exercises which ruled out any possibility to write up the PPB findings.

The fact that this study helped precipitate an interest in the Delphi technique and the requirement for new exercises is very gratifying and, in the opinion of the writer, all worthwhile. This, of course, may not be a consensus.

JW

CONTENTS

*Questionnaire 3 has been omitted due to its size. See Sections IV and VII for an explanation.

Report on an Experimental Use of The DELPHI Technique

I. INTENT

We had several objectives in mind--all experimental--when this study was originated. Initially, we wanted to understand the Delphi method, its mechanics and variations. Most importantly, we wanted to know if such a technique was applicable in the Agency and, if so, in what areas. Of nearly equal importance was the development of de-bugged and, if necessary, modified Delphi procedures which components could use as an aid in solving predictive, planning and policy problems. As a by-product, we hoped to obtain an approximation of the efficiency of Delphi, i.e., the amount of time and resources that are required to conduct exercises vis-a-vis the traditional conference approach.

Methodologically, we found ourselves confronted with a near vacuum as far as proven designs are concerned. A research of the basic studies by Dalkey [2, 4-8], Helmer [11, 12] and other RAND efforts gave us an appreciation for the state of Delphi "technology"; but no clear indication of an ideal or preferred

way to design and conduct an exercise. As a matter of fact [2] offered several exercise designs; however, it did not contain a precise statement correlating exercise objectives and design. In view of the foregoing, our hope was to try out a few of the available methods and gain an insight into exercise structures and applications through the use of a heuristic Delphi. From this experience we hoped to obtain de-bugged and reliable Delphi models which would be useful in real applications.

Substantively, our interest lay in assessing the relevancy of Delphi to Agency business with special emphasis on identifying application areas. To obtain this we needed a topic that was somewhat realistic and yet a "common denominator" for respondents.

Depending upon one's particular persuasion, a project such as this may be predestined to failure because of its scope, or predestined to success because any small degree of progress might be of value. In essence, the outcome of the exercise has in no way been spectacular. We do hope, however, that readers will agree with us that our results are partially successful, i.e., consensus was achieved for a large majority of the questions; the formal properties of the Delphi procedures were manifest and we did identify our mistakes for which subsequent exercises would not be penalized.

II. SUBJECT MATTER

Among the many subjects we might have chosen, or should have chosen, we somewhat arbitrarily selected the topic of Career Management. This subject does have some reasonable basis for selection and not just a natural curiosity or nosiness about personnel matters. The selection of the subject hinged on the background of the fifteen respondents. We felt, and still do, that respondents could, and would, offer comments on this subject even if their comments were based entirely on personal experience. Moreover, this subject conveniently fulfills the "common denominator" criterion.

One of the knottiest problems was our concern over the use of almanac type material. It is almost a tautology that if one wants to know something about factual data; one simply checks the appropriate reference material ... a Delphi isn't necessary. Conversely, if a forecasting or predicting exercise is conducted; we must, unfortunately, wait until year Y to see if event E occurred.

In order to meet our objectives with respect to method, we required material which would satisfy three conditions:

(1) We wanted questions where the respondents did not know the answer but had sufficient background information to make an informed estimate. (2) We wanted questions where there was a verifiable answer to check the performance of individuals, groups and, most importantly, the procedures. (3) We wanted questions with numerical answers so a relatively wide range of performance could be scaled, and where accuracy and error could be measured. As far as we could tell, almanac or general information material fits these criteria quite well.

As a consequence, we decided to do two things: include a section containing almanac questions and include a section on forecasting and evaluation; and thus have a two-part exercise. We recognize this structure may have caused some concern on the part of respondents and, furthermore, it may be a shortcoming of the study; but as things turned out there was more gain than loss from this approach.

III. RESPONDENT INFORMATION

Number of respondents 15

(14 men and 1 woman)

Experience (average in years)

Total	20
Agency	14
Other	6

Average Age 41

Low	26
High	56

Academic Background*

7	Business Administration and Economics
4	Social Science
3	Physical Science and Engineering
2	Law
1	Military Science

*Respondents cited more than one area of concentration.

IV. METHOD

Delphi is defined as a systematic set of procedures for eliciting and refining expert opinion. The procedures have three key features: (1) anonymous response and debate; (2) iteration and controlled feedback; (3) statistical group response. Instead of using the traditional approach toward achieving consensus through open discussion, Delphi eliminates committee activity altogether, thus reducing the influence of certain psychological factors, e.g., specious persuasion, the unwillingness to abandon publicly expressed opinions, group pressure for conformity (the bandwagon effect) and the role of the dominant individual. This technique replaces direct debate by a carefully designed program of sequential, individual, interrogations (questionnaires) interspersed with information and opinion feedback derived by computed consensus from earlier parts of the program [4, 10].

In general we adopted the basic features of the RAND design, but there were, however, two notable exceptions.

First of all, the exercise was truncated, i.e., three rounds. We believed numerical convergence of answers (consensus) would take place in three rounds and it would be of sufficient magnitude

to satisfactorily illustrate the procedures.* One of the main advantages of using four, five or more rounds is to enrich and refine the previously derived substantive arguments in support of various positions and estimates. In our case, this was not necessary as our exercise was not designed to provide decision-makers with profound substance which would serve as an aid in policy formulation.

Secondly, we were uncertain about the control of opinion feedback from round-to-round. Several factors contributed to this uncertainty, viz., the quantity of feedback (what is the right amount?), risk of scoring panel bias in editing narrative responses (particularly if respondent argumentation is one page per question) and a scale of measurement. As a consequence, we decided to vary the amount of feedback from round-to-round. In round 2, we asked the respondents to provide us with short statements in support of their estimates and/or evaluations. In round 3, we gave the respondents a full page for comment per question and fed back total (and unedited) information.

*This judgment is not wholly intuited. Dalkey provides a basis for this: [See Ref. 4, Section 7: Improvement with Iteration].

V. THE EXERCISE

Fifteen professionals, from various career services and directorates, served as volunteer-respondents. They answered a series of three questionnaires spread over several weeks. Typically, respondents had one week to answer each questionnaire. Altogether there were sixteen questions: nine sample questions on the key topic, Career Management, and seven non-related almanacs for calibration of method.

In round 1, respondents answered the questions relying on whatever background information they had at the time. We also requested respondents to rate themselves with respect to their individual confidence or competence to answer each question. A rating scale from 1-5 was used; a 5 reflected the highest confidence.

In questionnaire 2, the same familiar sixteen questions were fed back together with information on the Median and Inter-Quartile Range (IQR) of the first round responses (almanac questions only). Feedback on narrative or evaluative questions consisted of the number or ballots or votes for a particular position which was measured in terms of a binary response, i.e., yes-no; good-bad;

adequate-inadequate, etc. Self-rating averages (as a group) for each question were also included. Respondents were asked to revise their estimates giving the feedback data whatever weight they thought it deserved. For those individuals whose answers deviated markedly from the median, viz., outside the inter-quartile range, a short justification was required. For the evaluative questions one additional task was required, namely, a "Critique of Reasons," i.e., identification of reasons offered by other respondents which were assessed as unconvincing and a short statement as to why.

In round 3, respondents received the updated statistics and a summary of the justifications or argumentations for each question. Respondents were asked, again, to reappraise their answers in light of the new feedback information and then answer the questions for the last time. Section VI contains the summary statistics, balloting and voting on a round-by-round basis.

VI. RESULTS

The following section contains the final feedback and summary results.

22 December 1970

MEMORANDUM FOR: Participants in the PPB DELPHI Exercise

SUBJECT: Final Feedback of Results

1. The scoring panel would like to express their appreciation for your time and participation in the PPB DELPHI exercise. We were extremely gratified by: 1) the candor of the responses on all rounds, 2) the constructive criticisms which were offered to help improve the basic design and to shorten the time between rounds, and 3) the quality of the results--which were better than expected.

2. Since this was an experimental exercise, the results are not to be taken as valid for consideration by decision-makers. We cite the results as experimental, interesting and illustrative of the DELPHI method. Our project goal was to try out a technique, de-bug it and hopefully provide a tool suitable for use by management whenever they wished to obtain a consensus on some particular issue by using expert opinion. We think this goal has been achieved.

3. Our apologies for the delay in this final feedback questionnaire. Unfortunately, the delay could not be avoided. If there are any questions regarding the exercise, methods or results, the scoring panel will be available at your convenience.

[redacted]
OPPB

25X1

PPB DELPHI I

Final Feedback and Summary

Next 17 Page(s) In Document Exempt

Approved For Release 2005/04/18 : CIA-RDP82M00531R000400200002-1

VII. FINDINGS AND CONCLUSIONS

The Oracle at Delphi answered questions about the future but, unfortunately, the replies were garbled. We hope to do better here by staying away from prognostics and making no claims that we have found the small flames illuminating a clouded, dark and unexplored subject. But our Oracle is our exercise and its answers, no matter how fallible, should not be garbled. Empirically, several findings have emerged from the Delphi mist and discussion of each now follows:

1. Scope: The scope of Delphi is much broader than we first thought, i.e., technological forecasting of future events and developments. The empirical results indicate that the Delphi technique can be used in any context where it is appropriate to seek a consensus among experts on a particular subject. This is guardedly encouraging as exercises need not be restricted to technological probes.

2. Applications: In our view, the technique has great appeal from the standpoint of estimating, planning and even policy formulation. Clearly, this is not the same thing as saying Delphi has utility on all issues, occasions,

and under all circumstances. Like any management tool, the technique has certain limitations and there are data processing requirements which inhibit rapid multi-round solutions. But given normal circumstances and some computational support the procedures have a large potential in the aforementioned areas.

3. Candor of Responses: This was the most surprising (and refreshing) find of the lot. A certain amount of "leveling" was expected but nothing like the type encountered. This finding highlights the feature of anonymity and its importance should not be minimized. In private and post-exercise interviews with some respondents, they told us that the private and anonymous questionnaire provided the vehicle to express their true views on issues without putting their Fitness Report or career on the line. Moreover, the convenience of privately changing and modifying a position, in light of argumentation offered by fellow-respondents, was also noteworthy. Further, the questionnaire assures each respondent that his views will be surfaced and each counter-argument or criticism will have equal "air-time,"

thus, no respondent will be pre-empted. As one respondent put it, "unattributed answers are the keys to the Delphi chemistry."

4. Technical: Scientists are forever searching for more "inclusive" analyses of things they study. They are never so happy as when they have discovered a relationship between two kinds of phenomenon formerly considered independent. We had no such luck as we didn't discover anything new that RAND hadn't previously identified; however, there are a few technical factors which, in our view, deserve explicit mention and amplification, viz.:

(1) Careful attention to planning and designing an exercise is essential. This point is not stated here to add another tautology. Delphi questions must be carefully formulated for substance, form, simplicity, non-ambiguity and, most importantly, measurement. Actually, some form of pre-trial de-bugging should be employed if possible.

(2) With respect to measurement, we can't dodge the fact that our narrative-evaluative questions

were poorly scaled. We used the binary response form and, quite frankly, it just doesn't work.* What is needed is a modulus, a metric, or nominal scale of sufficient range, e.g., 0-10, 0-100%, to measure answers and their variation. Better still are measurement scales with semantic equivalents to permit numerical interpretation of qualitative factors.** To some extent we bailed out our lousy binary measurement scheme, on questions 8 and 15, in rounds 2 and 3 by going to the "desirability" scale which contains four degrees of qualification.

(3) Our exercise results, within certain limits, confirm or validate most of the functional relationships derived by Dalkey, et al. Specifically,

(a) The use of self-ratings as an index of group accuracy for any particular question, i.e., the higher the group rating the more accurate the group estimate and vice versa.

*If we may offer an opinion: We believe the technologist's criterion--Does it work? -- is it at least as effective in eliminating unwanted factors as the pure scientist's--Is it verified by laboratory experiment?

**Subsequent exercises have shown that this procedure works out quite well. See Refs. 14 & 15.

(b) Group size: 12 - 15 respondents is the approximate, minimum, number of experts needed per exercise. This may be a constraint to some managers on some issues when less than a dozen experts are available.

(c) Feedback: Questionnaires should be re-designed to allow respondents to synthesize their reasons for estimates. Reasons offered, per respondent, per question, should not exceed one paragraph. This is a trade-off between what can be managed by the scoring panel vs. bias and accuracy. The round 3 questionnaire (round 2 summary) was clearly too large and unwieldy, viz., 92 pages, the bulk of which was devoted to argumentation. The sheer size of this questionnaire prohibits its inclusion into this report. By reducing the size of the section: "Reasons For or Against," respondents will have to think a bit more and thereby synthesize their views--but isn't this what a manager wants?

VIII. REFERENCES

1. Aitchison, J., and Brown, J.A.C. The Log-Normal Distribution. Cambridge University Press, London, 1969.

2. Brown, B., Cochran, S., and Dalkey, N. The DELPHI Method II: Structure of Experiments. RAND, RM-5957-PR, June 1969.

3. Burrington, R.S., and May, D.C. Handbook of Probability and Statistics. Handbook Publishers, Inc., Sandusky, 1958.

4. Dalkey, N.C. The DELPHI Method (1): An Experimental Study of Group Opinion. RAND, RM-5888-PR, June 1969.

5. Dalkey, N. C., and Helmer, O. "An Experimental Application of the DELPHI Method to the Use of Experts, "Management Science, 9-456 (1963).

6. Dalkey, N.C., Brown, B., and Cochran, S. The DELPHI Method III: Use of Self-Ratings to Improve Group Estimates. RAND, RM-6115-PR, November 1965.

7. Dalkey, N.C. The DELPHI Method IV: Effect of Percentile Feedback and Feed-in of Relevant Facts. RAND, RM-6118-PR, February 1970.

8. Dalkey, N.C., Lewis, R., and Snyder, D. Measurement and Analysis of the Quality of Life: With Exploratory Illustrations of Applications to Career and Transportation Choices. RAND, RM-6228-DOT, August 1970.

9. Glushkov, V.M. "Forecasting Events on the Basis of Expert Opinion." Kibernetika, #2, pp 2-4, March-April 1969 (Translation in Soviet Cybernetics Review, 3 (10) 99-105 (October 1969).

10. Gordon, T. J., "New Approaches to DELPHI," in Technological Forecasting in Government and Industry, ed. by J. Bright, Prentice-Hall, 1968.

11. Gordon, T. J. and Helmer, O., Report on a Long-Range Forecasting Study, RAND, P-2982. With Brown, B., appendix to long-range forecasting study. September 1964.

12. Helmer, O. "Analysis of the Future: The DELPHI Method," in Technological Forecasting in Government and Industry, ed. by J. Bright. Prentice-Hall 1968.

13. Paxson, Edwin W., A DELPHI Examination of Divil Defense: 1. Questions, Issues, and Arguments; 2. Tabulation and Analysis of Responses. RAND, RM-6247/1&2, May 1970.

14. Waters, J., Clarke-DELPHI I: An Illustrative and Select Sample of Questions Re: Career Management in Government. Unpublished Report. American University. December 1970.

15. Wood, R., Waters, J., and Southern, J. INTEL DEL I. (S) Unpublished Report. Information Science Center, DIS. March 1971.

16. Wood R. (ed.) INTEL DEL II (C), Unpublished Report. Information Science Center, DIS. February 1971.